DOCUMENT RESUME

ED 265 223 TM 860 090

AUTHOR Baker, Eva L.; And Others
TITLE NIE Support to Assist in the Completion of USA Participation in the International Study of Written Composition. Final Report.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Dec 85
GRANT NIE-G-85-0006
NOTE 114p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Analysis of Variance; Cross Cultural Studies; Elementary Secondary Education; Hypothesis Testing; Models; *National Surveys; *Questionnaires; Research Methodology; *Sampling; *Scoring; Tables (Data); Writing (Composition); *Writing Evaluation; *Writing Skills
IDENTIFIERS *International Assn Evaluation Educ Achievement; National Center for Educational Statistics; National Institute of Education

ABSTRACT

This is a final report of support received through the National Institute of Education (NIE) to assist in research on the writing ability of United States children. The study began under the auspices of the International Association for the Evaluation of Educational Achievement (IEA) in order to contribute to a cross national assessment of students' writing proficiency. Aims of the study were: (1) to provide a conceptualization of the domain of writing; (2) to explore and describe various aspects of teaching written composition; (3) to describe how students in different countries respond to assignments; and (4) to test hypotheses about factors assumed to affect performance in written composition. A model was developed for the study which included three populations (representing three levels of schooling) and a series of instruments to be used. Research was organized around: (1) sampling of participant schools; (2) describing replacement strategy and procedure; and (3) describing the procedure for selecting classes. The remainder of the research addressed aspects of the sampling plan and procedure which were unique to the three individual populations of the model. Results of preliminary data analysis are presented in tables. A supporting document, Task 1 Training and Scoring Materials, is provided in an appendix. (LMO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED265223

FINAL REPORT ON NIE SUPPORT TO ASSIST IN THE COMPLETION OF USA PARTICIPATION IN THE INTERNATIONAL STUDY OF WRITTEN COMPOSITION

EVA L. BAKER
PRINCIPAL INVESTIGATOR

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. Griffith

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Grant Number
NIE-G-85-0006

TM 860 090

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California, Los Angeles

December 1985

The project presented or reported herein was in part performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

FINAL REPORT ON NIE SUPPORT TO ASSIST IN THE COMPLETION OF USA PARTICIPATION IN THE INTERNATIONAL STUDY OF WRITTEN COMPOSITION

UCLA CENTER FOR THE STUDY OF EVALUATION

Eva L. Baker
with the assistance of
Joan Herman, Beverly Cabello, Edys Quellmalz, and Anna Long

Grant # NIE-G-85-0006

Introduction

This constitutes a final report of the support received through the National Institute of Education (through November, 1985) to assist in research on the writing ability of United States children. The study, begun under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), is to contribute to a cross national assessment of students' writing proficiency. The research base for this effort derived from earlier CSE activities supported through NIE funds (from 1976-1981), and the design support for this endeavor was provided by the Spencer Foundation.

Considerable difficulty was experienced in trying to consolidate research support to permit this study to go forward. The MacArthur Foundation provided critical support to permit the data collection and to defray some costs of scoring. The NIE support assisted us in training raters, scoring, and initial data cleaning and analysis costs. Therefore, this report describes efforts supported by the NIE, but is not a final report of the study itself. We are in continued negotiations to secure funds for the full data analysis of this study.

This report will describe the background of the study briefly, but its focus will be on the training and scoring of the student essays. Preliminary data analyses will be reported, but we expect these to be rechecked, and to conduct additional analyses, before we forward them on to the IEA or publish a report on the U.S. National Study.

Background

Written composition skills is one of the clear school-based learning tasks on which there is almost universal agreement. Yet, the quality of students' writing in the USA is continually judged poor, whether the feedback comes from college admissions testing or the world of employment. Testimony to the difficulties in teaching writing can be inferred by the rapid expansion of remedial programs offered at institutions of higher learning to bring entering students' writing skills up to acceptable levels. And college-bound students are thought to perform at the highest levels in writing of all American secondary school students.

State departments of education have responded, as has the National Assessment of Educational Progress (NAEP), with measures designed to assess writing. Until recently, these measures were often "indirect", that is, they were made up of multiple choice items which correlated at moderate levels with actual writing performance on essays. In the course of the minimum competency testing movement, NAEP, a majority of states, and many local school districts have developed their own assessments of student composition using actual paragraphs or themes prepared by students.

While much of this effort is commendable, a good deal of it is flawed. First, the minimalist focus in some testing programs results in

students writing on tasks or topics where bare convention rather than thinking and organizing is the major outcome. For instance, many districts and states have at one time used a "letter" task and have emphasized letter format, e.g., location of the salutation, rather than letter content. Furthermore, many tasks are artificial, have no well developed audience (other than the teacher) or ask students to write about topics about which they may be drastically uninformed.

On psychometric grounds, many of these writing assessments are also inadequate. Because most assessment was developed for the purpose of individual placement, e.g., who to exempt from remedial English, underlying psychometric models focussed on rankings of students. Now, however, the explicit purposes of assessment have shifted to improving individual and group performance, rendering these normative models inappropriate. In fact, new psychometric models and validation strategies are available, in part developed with resources of the National Institute of Education. They are being applied in a subset of states, but should be more widely distributed.

Another serious issue in writing assessment is scoring or rating strategies. Because of the importance of improving student performance, the strategy recommended in this study, and being implemented in the larger international plan, provides diagnostic information. This strategy features an analytic scoring rubric, a guide that provides sub-scores on writing features. Analytic scoring has in the past emphasized only sentence-level flaws, e.g., spelling, but the guides proposed herein employ subscales designed to assess maturity and depth of ideas, sentence,

paragraph and theme organization, and style and audience, in addition to skills such as punctuation, syntactic appropriateness, and spelling. From such refined information, teachers and other instructional program planners can determine what aspects of instruction need modification and develop targeted rather than more costly global revisions of curricula.

An additional concern present in the conduct of most written composition assessment efforts is the cost of scoring. An initial investment was made wherein the composition raters are trained to a high level of agreement (and there is ample evidence such agreement is possible). The use of these scoring procedures reduced the amount of time spent on any one composition (four or five minutes for the most complex), and most compositions need to be scored by only one rater. An important side effect of such training (where it has been studied in laboratory like settings in school districts) is that the teachers so trained give students more writing assignments and, subsequently, writing performance improves.

A final concern with writing assessment involves the comparability of assessments. Does students' performance wobble annually depending upon the topic, the mode of discourse, e.g., exposition vs. narrative, or the scorers' orientation? Implemented in this study are procedures to permit the development of sets of task examples as well as methods to determine the extent to which performance <u>within</u> students differs depending upon the writing tasks.

The study was designed, then, to address all of the concerns above. It was further refined and improved to reflect the issues and experiences of other scholars in different countries. In particular, a developmental

sequence of writing tasks was proposed and, to a reasonable extent, will be ultimately tested by performance drawn from the three grade levels described. The actual concepts, objectives, procedural tasks, and measures are described in later portions of this report.

Background to the USA Study

The International Study was conceived by Alan Purves, at the University of Illinois. Early in 1980 he and staff of the UCLA Center for the Study of Evaluation collaborated on a proposal to be submitted to IEA for approval. Purves brought to the task understanding of the international community, of English language curriculum, and of the general products that might be usefully developed. Eva Baker, Edys Quellmalz, and Frank Capell, then all of CSE at UCLA, provided technical support in the design of domain-referenced writing specifications, alternative analytic rating research, descriptive studies of rater training and rater scoring stability, and psychometric analyses of writing. Merging these two perspectives was not automatic but made easier because of the joint agreement on the importance of the study, prior collaboration between UCLA and the University of Illinois, and the backgrounds of UCLA participants in the teaching of English.

As the proposed activity became more specific, with the full participation of an International Steering Committee, the UCLA role became focussed on the USA Study. Originally, funds for this study were thought to be available through other governmental sources. Subsequently, a self-supporting model, in which 8-10 states would serve as replicates, was planned. Unfortunately, this plan did not truly satisfy the IEA need for

national representation, and furthermore, many states had slightly different views of which writing tasks were important. (An important side-effect of this effort is that conceptually comparable writing assessment activities, reflecting a subset of the tasks proposed here, have been undertaken in Maryland, Illinois, Connecticut, and California.)

In the sections which follow, the history of the development effort is summarized and an overview of the IEA Study of Written Composition is provided.

To date, the principal support for the study came from The Spencer and The MacArthur Foundations and the NIE, with additional resources from the University of Illinois, the Center for the Study of Evaluation, UCLA, NIE, and The National Council of Teachers of English, among others. Accomplishments to date include:

1. An overall conceptual framework was developed.
2. A study design was developed, approved and implemented. All tasks were developed and piloted.
3. All questionnaires were developed and piloted.
4. Scoring guides were developed.
5. International scorer training was completed.
6. Sampling was conducted with assistance from the National Center for Educational Statistics.
7. Support was secured from the national government, educational policy makers, professional organizations, and leading individuals in measurement, evaluation, and English language teaching, and most significantly, from the Council of Chief State School Officers.
8. School contacts, data administration, and data return were accomplished.
9. Raters were trained.
10. Scoring of tasks was accomplished.
11. Preliminary data analysis of outcomes was conducted.

Overview of the IEA Study of Achievement in Written Composition

The teaching of written composition is a good example of a research area where both theoretical and empirical work is very much needed. The domain of writing is not well defined and, therefore, test construction is also at an early stage of development. In addition, there has been clear-cut agreement neither on the criteria of good writing nor on the methods of scoring. In this respect the Study of Written Composition differs clearly, for instance, from the Second Mathematics Study or the Second Science Study. During the first International Study Committee meeting it also became evident that the content and methods of instruction in written composition are to some extent unknown even in participating countries themselves.

Aims of the Study

The foregoing discussion of the current situation suggests that an important aim of the study is to provide a conceptualization of the domain of writing in general and of school writing in particular. This is a necessary step for the construction of a set of writing tasks which can be justified both theoretically and in terms of curricular validity.

The foregoing discussion also suggests that another important aim of the study is the exploration and description of (a) what is being taught in the instruction of written composition, (b) how written composition is being taught, (c) what kinds of exercises and assignments students are given in tests and examinations, (d) how much time is devoted to written composition, and (e) what characteristics are valued and what criteria are used in assessing performance in written composition.

A third major aim of the study is to describe how students in different countries respond to the assignments. This will include an attempt to score compositions using an internationally agreed upon scoring scheme as well as a more qualitative description of general patterns and variations in repsonse that might be associated with certain schools of thought and/or certain cultural variations.

A fourth aim of the study is to test some hypotheses about factors that are assumed to affect performance in written composition. These are based on a model developed for the study (see Figure 1).

Research Design

Populations and Samples

The definitions of the populations are:

Population A: Students at or near the end of primary education and the self-contained classroom.

Population B: Students at or near the end of comprehensive education, i.e., students who are in the last year of the shortest secondary program and those in longer programs who have completed the same number of years of schooling whether or not they have finished their program.

Population C: Students at or near the end of academic secondary school.

Each National Center will draw a representative stratified sample of students using strata that are important from the national point of view.

International Instruments

Data for the study was collected by means of a series of instruments specially prepared for the study -- writing tasks, questionnaires, and some attitude measures. Teachers administered all instruments. All writing

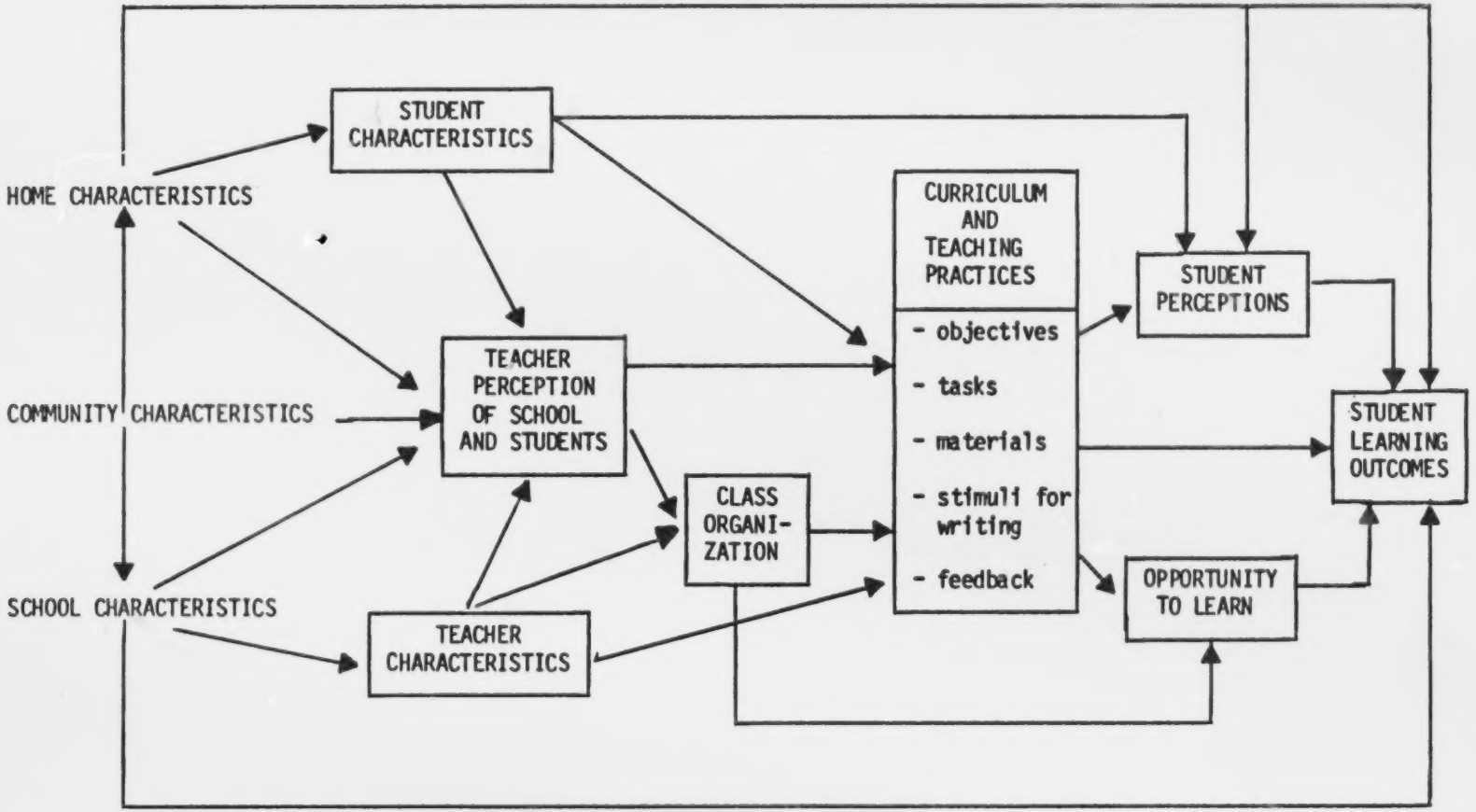

Figure 1. General Model of the Study (Main Constructs)

tasks were direct writing samples, in other words, no indirect "objective" tasks were used in the international component. Tasks were chosen on the basis of our model for the specification of the domain of writing and the information provided by participating countries on curricula, typical topics and writing tasks used in participating countries, and information about and samples of examinations.

There were some common writing tasks across the three populations (see Table 1). Tasks were rotated in order to cover a wide area of appropriate writing tasks without extending testing time too much.

National background information was obtained by means of the Curriculum Questionnaire, the Interview Schedule, National Context Questionnaire, and National Case Study Questionnaire.

Attitude questions were used to measure students' attitudes towards school and writing.

The Teacher Questionnaire was given to all the teachers who teach written composition to the students in the samples, to obtain information on teachers' qualifications, experience, teaching and feedback methods, etc.

The School Questionnaire was answered by the school principals and will ultimately provide data on the community and the school.

As a general principle, countries were also encouraged to add other types of national options to the basic international components, provided that the latter was not jeopardized, so that the total testing program in any given country (1) was perceived as appropriate by students, teachers and other interested parties, and (2) provided information that was

relevant in terms of national problems in the instruction of written composition. Because of cost constraints, no U.S. national options were exercised.

Sampling of Participating Schools

Of critical concern in this study was assuring a representative sample of respondents. This problem was complicated because the request for student time required two days of participation, far more than the usual "scientific survey." Furthermore, because participation was a local option, we were sensitive to the relatively low response rates achieved by studies of this sort. Third, because of the delay in funding for this effort, we would not have the leisure to engage in extensive follow-up or resampling. Therefore, our sampling strategy included leadership, practical, and technical concerns.

Leadership incentives. We believed that it was imperative that the schools be willing to participate. We did not have sufficient funds for lengthy, persuasive conversations. We secured the participation of the Council of Chief State School Officers whose leadership lent strong support to the idea of international comparisons and was also committed to improving the database of national indicators of educational quality. The timing of our request, following as it did national reports on USA educational performance, was fortuitous. We were able to transmit to each Chief Officer of State Education Departments a letter that described the study, the list of sampled districts in each particular state, the UCLA request letter to the district and schools for participation, and a draft letter to be sent with the Chief's signature, urging schools in our sample to participate. We believe that this step of securing the support of state leadership was critical to the success of the study.

Practical incentives. On the practical level, we also felt it essential to provide the local schools with motives to devote valuable instructional time to this study. In our letter, we stated our intentions to provide school level data back to the sending schools, both in terms of performance on each of the writing tasks and also with regard to national and international performance levels (when they were available). We focussed our efforts on the improvement of writing ability and also intended to provide analyses related to what particular instructional implications might be drawn from their school's results. We also promised to provide models for developing similar writing tasks and procedures for scoring writing according to the standards used in the study. In addition, we believe that we were assisted by the perception of writing as an important educational goal. We believe participation was forthcoming because school personnel could see the educational value of asking students to write, the major focus of our study.

Technical Concerns

In order to achieve the three objectives of the IEA study, as stated above, it was necessary to draw a sample which would represent the entire country for each of the specific populations. A variety of alternative approaches were considered to accomplish this representativeness; however, previous experience with the IEA Study of Science and available information on the total populations dictated the final approach. The remainder of this section discusses our approach, organized around three topics. The initial topic is the sampling of participant schools. This is followed by a description of the replacement strategy and procedure. And finally, a description of the procedure for selecting classes is presented.

The Selection of Schools

The approach chosen for the study was to draw a stratified probability-proportional-to-size (PPS) sample for each of the three populations. This approach produces a random sample that is representative of the population and allows for estimation of national proficiency levels. In order to implement the sampling, it was necessary to obtain information on the total U.S. populations of schools. This information was provided by the National Center for Educational Statistics*(NCES) which allowed access to their data base describing public and private schools within the United States. It should be noted that the available information on the public and private schools differed both in its recency and its content. The public school information was derived from a survey current through the 1982 school year while the private school information was current through 1980. However, it was felt that the difference in recency and type of information could be minimized by the design of the sample and sampling procedure.

In drawing a stratified PPS sample for each population, the stratification dimension selected was the type of school (i.e., public or private). Other potential dimensions for stratification such as community type, socio-economic status, and student racial/ethnic composition were also considered but discarded. Experience with previous IEA studies had indicated that, if the sample size is sufficiently large, these dimensions will be adequately represented in a sample drawn through a simple random

*We would like to express our appreciation to NCES, and particularly to Dr. Larry Suter, for their cooperation and assistance with this aspect of the study.

selection procedure. Since the smallest sample for the current study was over a 100 schools, it was felt that reliance on random sampling by school type was sufficient. The choice of this approach was also consistent with the available information at NCES where the selection was to be performed.

Procedurally, the sampling required the creation of a computer routine which would enter the appropriate data base, compute a measure of size for each school, order the schools with respect to this measure, and then select every "Nth" school as a sample participant. The measure of size used in the sampling was an estimate of the number of students enrolled in a school at the target grade level. The estimate was derived by dividing the reported school enrollment by the reported number of grade levels served by the school. This estimate assumes uniform distribution of students over grades which while potentially inaccurate in some cases was the most appropriate procedure for estimating size given the available data.

In addition to ordering the schools on size, schools were also ordered by state. While this, the secondary ordering did not represent a full stratification it increased the geographic coverage of the sample. Geographic coverage was viewed as desirable since it would enhance the likelihood of national representativeness, spread the respondent burden across states, and reduce the burden at the district level.

After schools were ordered on size and state, two additional pieces of information were required to perform the selection: a random start value and a constant interval size. The former was drawn from a table of random numbers. The latter was calculated for each stratum by dividing the sum of

the measures of size by the desired number of schools for the sample.* When these two pieces of information were provided to the computer routine**, a list of sample schools was produced. This procedure was repeated for each stratum and for all three populations.

Selection of Replacement Schools

In addition to drawing the primary sample for the study, it was necessary to anticipate the need for replacements should any of the participant schools decline to participate in the study. It was decided that the most efficient method for handling the replacements was to draw parallel samples. This approach was used succesfully in the IEA Science Study and decreased the handling and logistical problems involved with replacements. Since the timeline for the conduct of the U.S. study was exceedingly short, it was felt that these logistical constraints outweighed any other considerations. Thus, a total of four samples for each stratum within each population was drawn. This included a primary sample and three replacement waves.

Actual use of the replacement schools was made on a case by case basis. That is, the procedure for replacement was initiated only after a school refused participation and there were insufficient participating schools to satisfy the state quota derived from the primary sample. The state quota was set by the number of schools in the primary sample that originated from each state. When the need for replacement arose, schools

*The sum of the size measure was computed from previous analyses of the total school populations.

**The computer routine used for the sampling was a modified version of that used at NCES for the IEA science study.

from the same state and from the first available replacement wave were used as potential replacements. Final selection was performed by selecting the schools on the basis of the measure of size so that the replacement was as similar as possible to the non-participant.

It should be noted that prior to the selection of the four sample waves for each population certain exclusions were made from the total pool of schools contained in the NCES data base. Specifically, schools were excluded if they were institutions which served special non-normal populations of students, or if they were not operating in the most recently reported year, or if they were located in other than the 50 states and the District of Columbia. It was felt that these exclusions were consistent with the intent of the study and insured high data quality.

Selection of Classrooms

The last aspect of the general sampling approach was the specification of the particular classroom to participate in the study. The design of the international study called for one classroom to participate from each selected school at the target grade level. To facilitate identification of the classroom at the school level, it was decided that a simple but appropriate method of selection must be employed. Therefore, it was decided that the class should be selected on the basis of size. Since only one class was needed at each school, the school had only to identify its largest class from the appropriate grade as the participating class. This approach was consistent with the strategy chosen for the selection of schools and greatly minimized the logistical problems of class selection. Implementation of this process was left to the School Coordinators, although they were assisted by a description of the procedure in the Coordinator Manual provided to them.

To summarize, the general sampling approach addressed the issue of national representation while balancing constraints of available information and logistics. The resultant approach was to draw a stratified PPS sample from two strata in each of the three populations. The strata were defined as public schools and private schools. Within each stratum, schools were ordered by size and state prior to selection. Four parallel samples were drawn for each population and from within each stratum. This strategy allowed for easy replacement while insuring the same sample structure as in the primary sample. Classroom selection was also performed using size as the selection criteria. Since only a single class per school was needed, this approach resulted in the selection of the largest class at the target grade.

The remainder of this document addresses the aspects of the sampling plan and procedure which were unique to the three individual populations. Specifically, each of the following three sections provides a definition of the population and a specification of sample size.

Population A: Grade Six Schools

Definition of Target Population

Population A is defined as students at or near the end of primary education and in a self-contained classroon. For most states, this occurs during the sixth grade and the median age of the students is 11 years. Because, the end of compulsory schooling in each of the participating states is at age 16 years, virtually 100% of sixth grade students in the population attend school.

(1) Desired target population: all sixth grade students.

(2) Defined target population: all sixth grade students on the dates of testing.

(3) Excluded population: all students above or below Grade 6 and all students not in school on the dates of testing.

Sample Size

The international study design requires a sample of at least fifty classrooms for this grade level. The United States study was designed with slightly more than double this sample size thus providing greater precision for national estimates. Specifically, the Population A sample was designed to include 100 schools from the Public School stratum and 10 schools from the Private School stratum for a total of 110 grade six schools. The allocation of the total sample to the two strata was a function of interest and resources. It should be noted that a 25% oversampling of the Public School stratum and a 50% oversampling of the Private School stratum was actually performed from this population to help insure that at least the required number of schools would be available for participation. That is, since the information in the data base was dated it was anticipated that some percentage of the sampled schools would be ineligible for study participation (either because the school no longer existed, or because the school no longer served the grade level of interest). This procedure resulted in 125 public schools and 20 private schools in each sample wave. Our target was 110 schools.

Population B: Tenth Grade Schools

Definition of Target Population

Population B is defined by students who are at the end of compulsory schooling. The end of compulsory schooling in each of the participating states is at age 16 years. In general, students become 16 years old just after or during the last half of their tenth grade year. Thus, most tenth grade students are in the last year of compulsory education.

(1) Desired target population: all tenth grade students.

(2) Defined target population: all tenth grade students on the dates of testing.

(3) Excluded population: all students above or below Grade 10 and all students not in school on the dates of testing.

Sample Size

The international study design requires a sample of at least 100 classes at the tenth grade level. The United States study was designed for slightly more than double this number to a total of 220. The rationale for the larger sample was the greater precision to be achieved in the population estimates to be derived from the sample. The Population B sample was structured so that 200 public schools and 20 private schools would participate in the final study. As in Population A, a 25% oversampling of the Public School stratum and a 50% oversampling of the Private School stratum was performed. This resulted in 250 Public Schools and 40 Private Schools in each of the four sample waves.

Population C: Twelfth Grade Academic Students

Definition of Target Population

Population C is defined by students who are at the end of an academic program in secondary school. For the states participating in this study, this occurs in grade 12 and the mean age of 12th grade students is 17 years.

(1) Desired target population: all twelfth grade students in academic programs.

(2) Defined target population: all twelfth grade students in academic programs on the dates of testing.

(3) Excluded population: all students above or below Grade 12, all students not in school on the dates of testing, and all Grade 12 students not on academic tracks.

Sample Size

The international requirements for this population are the same as that of Population A, namely 50 classes. The United States study differs from that by slightly more than double. Specifically, a total of 110 schools was selected as the desirable number of participants for the U.S. study. The allocation of these across strata was the same as for Population A with 100 public schools and 10 private schools. It should be noted at this point that the schools selected for participation in the Population C study were limited to those schools which were participating in the 10th grade study. Thus, the final selection of schools for Population C was linked to the Population B sampling with that population (B) providing the pool of eligible schools for the Population C sample. As with the previous populations a 25% oversampling of the Public School stratum and a 50% oversampling of the Private School stratum was performed. This produced 125 public schools and 20 private schools in the four sample waves.

Response to Sample

The results of the sampling plan were encouraging. The IEA minimums for 6th grade were 50 classrooms, for 10th grade, 100 classrooms, and for 12th grade, 50 classrooms. Our intention was to double the minimum sample sizes, so our functional targets were 100 classrooms at 6th grade, 200 at 10th grade, and 100 at 12th grade, although our sampling procedures used 110, 220, and 110 as desired samples.

As noted earlier, we oversampled by the reported percentages to expedite the data collection process. In Table 1 below, we report the actual number of classrooms included in our scoring and preliminary analyses. Note that private schools are slightly over represented. These figures do

not include data from three classrooms which arrived too late to be scored.

Table 1

Classroom Data at Each Grade Level

SCHOOLS	6TH GRADE	10TH GRADE	12TH GRADE	TOTAL
PUBLIC	72	167	92	331
PRIVATE	8	26	11	45
Totals	80	193	103	376
Percent of Targets	80%	96+%	103%	94%

Twelfth grade classes were selected from schools with 10th grade participating classrooms. This decision was to facilitate (reduce) the number of school contacts required, and more importantly, to provide a school level factor for data analyses. Table 2 below, therefore presents the number of unique schools included in the study analyses.

Table 2

Total Unique Schools Providing Data

	PUBLIC	PRIVATE	TOTAL
POPULATION A: (6th grade only)	72	8	80
POPULATION B: (10th grade only)	75	15	90
POPULATION C: (10th and 12th grades)	92	11	103
TOTAL	239	34	273

Scoring and Training

Logistics for training raters and scoring data were enormously complex. Resources were not available to score all papers. After conferring with NCES and NIE personnel, it was agreed to score eight papers per task for each classroom. Actual papers scored vary because of student absences and differential class enrollments, and because of the complex rotation patterns for tasks, within classroom sampling by task occurred.

Scoring Approach

IEA scoring procedures were followed. Essentially these required scoring on three analytic subscales, organization, style, and content detail and appropriateness, and a general impression scoring. Task differences also influenced patterns of training and scoring. We scored similar tasks (1A-E, 6, 7) in the first scoring session held in June, 1985, and tasks 3, 5, 9 in the second session. In addition, 20% of the papers were scored by two separate raters (double scored). We also used check papers (20 per task per day) to assure that raters were staying on the scale. In addition, our plan required the scoring of calibration papers (to permit equating for international comparison). Calibration papers were scored by everyone for each task. Approximately 20 calibration papers were used for each of the 10 tasks.

Selection of Scoring Directors and Raters

A list of potential scoring directors and raters was devised based on our collective experience in the field of writing assessment. We also sought recommendations from districts throughout California as well as from other university institutions both within UCLA and at other universities.

Scoring directors were selected on the basis of their experience in using an analytic scoring scheme or in participating as a trainer in large scale writing assessment efforts. Raters were also selected on the basis of their experience as raters in large scale or district wide scoring efforts. However, it was not mandatory that they be familiar with the analytic types of scoring rubrics. Public and private school personnel formed the rating pool. For the June session, 40 raters were used; for the August session 20 raters were used.

Validation Scoring

A validation meeting for the June scoring session was held including two CSE coordinators and four scoring directors. This meeting had several purposes: 1) to acquaint the scoring directors with the IEA study and with the procedures for training scorers and scoring the writing samples; 2) to review the scoring rubric for each writing task to be scored and make any necessary clarifications; 3) to select and score training, qualifying and check papers.

In preparation for this meeting, the CSE coordinators selected a pool of papers for each task which represented various proficiency levels and scoring problems. These papers were selected from the larger sample to be scored. The validation meeting resulted in a pool of scored training, qualifying, and checkpapers, complete with scores and comments. These were later xeroxed and prepared for the purpose of training the raters.

Prior to the validation session notebooks were devised for the scoring directors. These contained a list of potential training papers, qualifying papers, and checkpapers as well as copies of each type of paper. A copy of the scoring manual and scoring sheets were also included. Each scoring

director was provided with a notebook containing the information corresponding to the task(s) which she was responsible for. After the validation similar notebooks were devised for the raters. These included a list and copies of the training papers and a list and copies of checkpapers for the first day.

Validation of training, qualifying and checkpapers for the August session was conducted via the mail and telephone. This was possible because the scoring directors as well as the raters were well acquainted with all aspects of the training and scoring session because of their experience in the June scoring session.

Rater Training and Scoring Sessions

Orientation. The orientation was conducted by the CSE Coordinators during the first morning of the session. Raters were introduced to the coordinators, scoring directors, and support staff. They received an overview of the IEA study and of the general training and scoring procedures.

Afterward the raters were assigned to their training and scoring group. In making the assignment, coordinators spread groups of raters who came from the same district across scoring groups. The intent here was to ensure that personnel coming from the same district would take with them the variety of scoring skills offered by the IEA training and scoring sessions.

After the overview and assignment of groups, the scoring directors began to train their raters on the interpretation and application of the IEA scoring rubric to the specific writing tasks and samples to be scored.

Rater training. After giving everyone the opportunity to carefully read the scoring criteria for a particular writing task, the trainers (using overhead transparencies) illustrated each score for each scoring element with a writing sample. After this had been done for all of the scoring criteria and range of scores for a particular task (for example, task 6), raters were given the opportunity informally to score a few (3-5) training papers. The trainers took a voice count and tallied the scores for the training papers to get a sense of how well everyone was applying the rubric. These scores also signaled potential misunderstandings or unusual aspects of the papers. The trainers and raters discussed these, giving examples to clarify the problems. The process of scoring training papers and discussing the scores continued until the trainers felt that the raters were ready to pass the qualifying papers. Agreement was defined as within one score point on the six point scale for each scale scored. The qualifying papers were regarded as a test to determine the extent to which individual raters were accurate, as well as the extent to which the entire group of raters (for a particular task) agreed with each other. If a rater, or the group as a whole, did not pass the qualifying set, they were retrained. To pass the test, both individually and as a group, the scores reflected an agreement of at least .80. For example, if there were 10 qualifying papers for task 6, then each rater was allowed only two errors for the entire set of papers across all of the scoring elements. In both the June and August sessions, all groups passed their qualifying sets on the first administration and no group retraining was necessary.

After passing the qualifying sets, raters went on to score the sample papers. However, to ensure that high reliability was maintained both individually and for the group, a set of approximately 20 checkpapers had to be scored on a daily basis; half of these were scored in the morning and half in the afternoon. Again, 80% agreement was the standard. As necessary, the individual raters were taken aside to discuss the particular errors they had made, and then retested.

Preliminary Data Analysis

Data analysis is presented for Tasks 1a, 1b, 1c, 1d, 1e, 6, and 7.

Initially, a total of 11,131 essays were scored for the above seven tasks. In this first cleaning run, 2.20 percent (245 cases) of these cases were lost due to missing or invalid identification numbers, reducing the number of cases to 10,886 for the seven tasks. In the final cleaning run, 6.23 percent of the data was lost to missing grade levels, population level (public vs. private), and missing task numbers (a total of 678 cases). The final data set contains 10,208 essays (see Table 3).

Table 3
Deletions from Data Set

	ESSAYS	PERCENT
Initial Essays Scored:	11,131	100.00
Deleted Due to ID's	- 245	2.20
Other Deleted Data	- 678	6.09
TOTAL DATA SET	10,208	91.71

Table 4 presents distributions of essays in final data set by task, grade, and public or private school population. Table 5 displays the total essays scored by task across grade levels.

Table 4

Essays in Final Data Set
by Task, Grade and Population

6th Grade

	1A	1B	1C	1D	1E	6	7
PUBLIC	355	301	293	370		559	
PRIVATE	26	27	27	27		39	
TOTAL	381	328	320	397		598	

10th Grade

	1A	1B	1C	1D	1E	6	7
PUBLIC	882	946	870		779	737	687
PRIVATE	149	152	156		141	143	131
TOTAL	1031	1098	1026		920	880	818

12th Grade

	1A	1B	1C	1D	1E	6	7
PUBLIC					576	815	747
PRIVATE					79	99	95
TOTAL					655	914	842

Table 5

Totals of Scored Essays by Task

TASK	1A	1B	1C	1D	1E	6	7	TOTAL
6th GRADE	381	328	320	397		598		2024
10th GRADE	1031	1098	1026		920	880	818	5773
12th GRADE					655	914	842	2411
TOTAL	1412	1426	1346	397	1577	2391	1660	10,208

Preliminary Descriptive Statistics

Descriptive statistics at the student level were computed by task and grade level. These analyses were computed for each of five outcome variables: OI - Overall Impression; COMP - a competency measure averaging three analytic scores; CONTENT - Content appropriateness and specificity; ORGAN - Organization; and STYLE. The scores ranged from 1-6. These are presented in the Tables 6 - 18 below.

Table 6

Means and Standard Deviations for Task 1 by Task and Grade
Task = 1A Grade = 6

VARIABLE	N	MEAN	STD DEV
OI	380	2.765	0.907
COMP	380	2.807	0.831
CONTENT	380	2.702	0.943
ORGAN	380	2.936	0.908
STYLE	380	2.781	0.931

Table 7

Means and Standard Deviations for Task 1 by Task and Grade
Task = 1A Grade = 10

VARIABLE	N	MEAN	STD DEV
OI	1035	3.296	0.952
COMP	1035	3.275	0.885
CONTENT	1035	3.283	1.071
ORGAN	1035	3.339	0.903
STYLE	1035	3.202	1.033

Table 8

Means and Standard Deviations for Task 1 by Task and Grade
Task = 1B Grade = 6

VARIABLE	N	MEAN	STD DEV
OI	328	2.896	1.194
COMP	328	2.970	1.122
CONTENT	328	2.817	1.257
ORGAN	328	3.146	1.129
STYLE	328	2.948	1.181

Table 9

Means and Standard Deviations for Task 1 by Task and Grade
Task = 1B Grade = 10

VARIABLE	N	MEAN	STD DEV
OI	1104	3.411	0.917
COMP	1104	3.440	0.854
CONTENT	1104	3.411	1.034
ORGAN	1104	3.501	0.872
STYLE	1104	3.409	0.924

Table 10

Means and Standard Deviations for Task 1 by Task and Grade
Task = 1C Grade = 6

VARIABLE	N	MEAN	STD DEV
OI	320	3.003	0.990
COMP	320	3.004	0.953
CONTENT	320	2.956	1.058
ORGAN	320	3.121	0.976
STYLE	320	2.934	0.037

Table 11

Means and Standard Deviations for Task 1 by Task and Grade
Task = 1C Grade = 10

VARIABLE	N	MEAN	STD DEV
OI	1031	3.491	0.876
COMP	1030	3.497	0.851
CONTENT	1031	3.465	0.999
ORGAN	1031	3.568	0.879
STYLE	1030	3.457	0.933

Table 12

Means and Standard Deviations for Task 1 by Task and Grade
Task = 1E Grade = 10

VARIABLE	N	MEAN	STD DEV
OI	927	2.881	0.917
COMP	927	2.800	0.827
CONTENT	927	2.555	0.949
ORGAN	927	2.907	0.920
STYLE	927	2.938	0.887

Table 13

Means and Standard Deviations for Task 1 by Task and Grade
Task = 1E Grade = 12

VARIABLE	N	MEAN	STD DEV
OI	656	3.135	0.876
COMP	655	2.986	0.779
CONTENT	656	2.711	0.938
ORGAN	655	3.068	0.898
STYLE	655	3.180	0.800

Table 14

Means and Standard Deviations for Task 6 by Grade
Grade = 6

VARIABLE	N	MEAN	STD DEV
OI	705	2.561	1.624
COMP	664	2.382	1.168
CONTENT	664	2.296	1.232
ORGAN	664	2.246	1.209
STYLE	664	2.603	1.218

Table 15

Means and Standard Deviations for Task 6 by Grade
Grade = 10

VARIABLE	N	MEAN	STD DEV
OI	1150	3.099	0.959
COMP	1143	3.085	0.815
CONTENT	1144	3.124	0.968
ORGAN	1143	2.928	0.922
STYLE	1144	3.202	0.805

Table 16

Means and Standard Deviations for Task 6 by Grade
Grade = 12

VARIABLE	N	MEAN	STD DEV
OI	1046	3.530	0.986
COMP	1039	3.474	0.848
CONTENT	1039	3.527	0.987
ORGAN	1039	3.390	0.963
STYLE	1039	3.505	0.836

Table 17

Means and Standard Deviations for Task 7 by Grade
Grade = 10

VARIABLE	N	MEAN	STD DEV
OI	1045	3.039	1.011
COMP	1036	3.032	0.832
CONTENT	1036	3.031	0.981
ORGAN	1036	2.938	0.954
STYLE	1036	3.127	0.799

Table 18

Means and Standard Deviations for Task 7 by Grade
Grade = 12

VARIABLE	N	MEAN	STD DEV
OI	1014	3.339	0.985
COMP	1005	3.308	0.815
CONTENT	1005	3.325	0.975
ORGAN	1005	3.241	0.919
STYLE	1005	3.357	0.792

Analyses of variance were computed for those tasks assessed at more than one grade level. These analyses are presented in Table 19 below.

Table 19

SAS
TASK=1A

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN BY GROUP = 1420

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1416 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1A

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: G1

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	78.34442570	39.17221285	44.28	0.0001	0.058975	29.822
ERROR	1413	1250.09342741	0.88470872		ROOT MSE		G1 MEAN
CORRECTED TOTAL	1415	1328.43785311			0.94058956		3.1539548

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	78.34442570	44.28	0.0001

BEST COPY

SAS
TASK=13

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	A B

NUMBER OF OBSERVATIONS IN BY GROUP = 1436

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1432 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=18

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: GI

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	67.03930806	67.03930806	68.68	0.0001	0.045829	29.999.
ERROR	1430	1395.77633439	0.97606737		ROOT MSE		GI MEAN
CORRECTED TOTAL	1431	1462.81564246			0.98796122		3.29329809

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	67.03930806	68.68	0.0001

BEST COPY

SAS
TASK=1C

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	A B

NUMBER OF OBSERVATIONS IN BY GROUP = 1354

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1351 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1C

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: GI

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	58.30617817	58.30617817	71.20	0.0001	0.050135	26.804
ERROR	1349	1104.67679741	0.81888569		ROOT MSE		GI MEAN
CORRECTED TOTAL	1350	1162.98297557			0.90492303		3.3760177

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	58.30617817	71.20	0.0001

BEST COPY

SAS
TASK=1E

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN BY GROUP = 1674

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1588 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1E

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: DI

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	32.18586615	16.09293307	19.67	0.0001	0.024221	30.244/
ERROR	1585	1296.67244619	0.81808987		ROOT MSE		DI MEA.
CORRECTED TOTAL	1587	1328.85831234			0.90448321		2.9905541.

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	32.18586615	19.67	0.0001

BEST COPY

SAS
TASK=1A

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN BY GROUP = 1420

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1416 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: COMP

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	61.11656706	30.55828353	40.23	0.0001	0.053873	27.6772
ERROR	1413	1073.33093765	0.75961142		ROOT MSE		COMP MEAN
CORRECTED TOTAL	1415	1134.44750471			0.87155089		3.1490113

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	61.11656706	40.23	0.0001

BEST COPY

SAS
TASK=18

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	A B

NUMBER OF OBSERVATIONS IN BY GROUP = 1436

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1432 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=18

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: COMP

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	55.92893639	55.92893639	65.65	0.0001	0.043896	27.6911
ERROR	1430	1218.18209713	0.85187559		ROOT MSE		COMP MEAN
CORRECTED TOTAL	1431	1274.11103352			0.92297107		3.3331005

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	55.92893639	65.65	0.0001

BEST COPY

SAS
TASK=IC

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	A B

NUMBER OF OBSERVATIONS IN BY GROUP = 1354

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1350 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=IC

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: COMP

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	59.47676735	59.47676735	77.31	0.0001	0.054241	25.9441
ERROR	1348	1037.04471413	0.76932100		ROOT MSE		COMP MEAN
CORRECTED TOTAL	1349	1096.52148148			0.87710946		3.3807407

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	59.47676735	77.31	0.0001

BEST COPY

SAS
TASK=1E

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN BY GROUP = 1674

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1567 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1E

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: COMP

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	23.88907836	11.94453918	18.13	0.0001	0.022376	28.1044
ERROR	1564	1043.74293173	0.65892862		ROOT MSE		COMP MEAN
CORRECTED TOTAL	1566	1067.63201008			0.81174418		2.88216761

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	23.88907836	18.13	0.0001

BEST COPY

SAS
TASK=1A

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN BY GROUP = 1420

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1416 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1A

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: ORGAN

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	46.49689977	23.24844988	28.41	0.0001	0.038653	28.006.
ERROR	1413	1156.44942792	0.81843555		ROOT MSE		ORGAN MEAN
CORRECTED TOTAL	1415	1202.94632768			0.90467428		3.2302259.

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	46.49689977	28.41	0.0001

BEST COPY

SAS
TASK=1B

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	A B

NUMBER OF OBSERVATIONS IN BY GROUP = 1436

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1432 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1B

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: ORGAN

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	31.95259444	31.95259444	36.35	0.0001	0.024790	27.410
ERROR	1430	1256.97198657	0.87900139		ROOT MSE		ORGAN MEAN
CORRECTED TOTAL	1431	1288.92458101			0.93755074		3.4203510

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	31.95259444	36.35	0.0001

BEST COPY

SAS
TASK=1C

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	A B

NUMBER OF OBSERVATIONS IN BY GROUP = 1354

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1351 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1C

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: ORGAN

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	48.68625425	48.68625425	59.64	0.0001	0.042341	26.092
ERROR	1349	1101.17006996	0.81629064		ROOT MSE		ORGAN MEAN
CORRECTED TOTAL	1350	1149.85632420			0.90348804		3.4626202

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	48.68625425	59.64	0.0001

BEST COPY

SAS
TASK=1E

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN BY GROUP = 1674

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1587 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1E

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: ORGAN

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	23.10382770	11.59191385	13.90	0.0001	0.017246	30.654
ERROR	1584	1321.12997192	0.83404670		ROOT MSE		ORGAN MEAN
CORRECTED TOTAL	1586	1344.31379962			0.91326157		2.9792060

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	23.10382770	13.90	0.0001

BEST COPY

SAS
TASK=1A

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN BY GROUP = 1420

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1416 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1A

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: CONTENT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	93.66716937	46.83358468	43.44	0.0001	0.057922	33.2044
ERROR	1413	1523.45147470	1.07816806		ROOT MSE		CONTENT MEAN
CORRECTED TOTAL	1415	1617.11864407			1.03834872		3.1271186

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	93.66716937	43.44	0.0001

BEST COPY

SAS
TASK=18

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	A B

NUMBER OF OBSERVATIONS IN BY GROUP = 1436

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1432 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=18

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: CONTENT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	89.26985719	89.26985719	75.25	0.0001	0.049994	33.255
ERROR	1430	1696.32511488	1.18624134		ROOT MSE		CONTENT MEAN
CORRECTED TOTAL	1431	1785.59497207			1.08914707		3.2751396

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	89.26985719	75.25	0.0001

BEST COPY

SAS
TASK=1C

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	A B

NUMBER OF OBSERVATIONS IN BY GROUP = 1350

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1351 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1C

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: CONTENT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	63.34762525	63.34762525	61.66	0.0001	0.043710	30.3023
ERROR	1349	1385.91514306	1.02736482		ROOT MSE		CONTENT MEAN
CORRECTED TOTAL	1350	1449.26276832			1.01359007		3.34492963

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	63.34762525	61.66	0.0001

BEST COPY

SAS
TASK=1E

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN BY GROUP = 1574

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1588 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1E

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: CONTENT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	25.17493814	12.58746907	14.00	0.0001	0.017364	36.103%
ERROR	1585	1424.63614499	0.89882407		ROOT MSE		CONTENT MEAN
CORRECTED TOTAL	1587	1449.81108312			0.94806332		2.6259445

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	25.17493814	14.00	0.0001

BEST COPY

SAS
TASK=1A

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN BY GROUP = 1420

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1416 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1A

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: STYLE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	49.34710630	24.67355315	24.34	0.0001	0.033306	32.585
ERROR	1413	1432.26235698	1.01363224		ROOT MSE		STYLE MEAN
CORRECTED TOTAL	1415	1481.60946328			1.00679305		3.0896892

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	49.34710630	24.34	0.0001

BEST COPY

SAS
TASK=18

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	A B

NUMBER OF OBSERVATIONS IN BY GROUP = 1436

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1432 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=18

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: STYLE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	53.79870545	53.79870545	54.99	0.0001	0.037030	29.939
ERROR	1430	1399.06093142	0.97836429		ROOT MSE		STYLE MEA
CORRECTED TOTAL	1431	1452.85963687			0.98912299		3.3037709

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	53.79870545	54.99	0.0001

BEST COPY

SAS
TASK=1C

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	A U

NUMBER OF OBSERVATIONS IN BY GROUP = 1350

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1350 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1C

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: STYLE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	66.79773665	66.79773665	72.50	0.0001	0.051030	28.787
ERROR	1348	1241.2422o335	0.92080287		ROOT MSE		STYLE MEAN
CORRECTED TOTAL	1349	130d.00000000			0.95958474		3.3333333

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	66.79773665	72.50	0.0001

BEST COPY

SAS
TASK=1E

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN BY GROUP = 1674

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 1587 OBSERVATIONS IN BY GROUP CAN BE USED IN THIS ANALYSIS.

SAS
TASK=1E

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: STYLE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	27.01806812	13.50903406	18.41	0.0001	0.022718	28.162+
ERROR	1584	1162.23713036	0.73373556		ROOT MSE		STYLE MEA
CORRECTED TOTAL	1586	1189.25519849			0.85658366		3.04158790

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	27.01806812	18.41	0.0001

BEST COPY

BEST COPY

SAS

TASK 6 ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN DATA SET = 2472

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2408 OBSERVATIONS IN DATA SET CAN BE USED IN THIS ANALYSIS.

SAS 12:50 WEDNESDAY, NOVEMBER 27, 1985

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: OI

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V
MODEL	2	332.87751902	166.43875951	122.99	0.0001	0.092786	37.17
ERROR	2405	3254.69723181	1.35330446		ROOT MSE		OI MEA
CORRECTED TOTAL	2407	3587.57475083			1.16331615		3.129568

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	332.87751902	122.99	0.0001

SAS

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN DATA SET = 2472

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2358 OBSERVATIONS IN DATA SET CAN BE USED IN THIS ANALYSIS.

SAS

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: COMP

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V
MODEL	2	415.88125499	207.94062749	253.36	0.0001	0.177069	29.58
ERROR	2355	1932.81462665	0.82072808		ROOT MSE		COMP MEA
CORRECTED TOTAL	2357	2348.69588163			0.90594044		3.0650266

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	415.88125499	253.36	0.0001

SAS

BEST COPY

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN DATA SET = 2472

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2359

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	415.88125499	283.36	0.0001

1 SAS

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN DATA SET = 2472

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2359 OBSERVATIONS IN DATA SET CAN BE USED IN THIS ANALYSIS.

1 SAS 12:50 WEDNESDAY, NOVEMBER 27, 1985

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: STYLE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V
MODEL	2	278.22610469	139.11305235	165.78	0.0001	0.123366	28.836
ERROR	2356	1977.06088132	0.83915967		ROOT MSE		STYLE MEA
CORRECTED TOTAL	2358	2255.28698601			0.91605675		3.1767698

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	278.22610469	165.78	0.0001

1 SAS

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN DATA SET = 2472

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2358 OBSERVATIONS IN DATA SET CAN BE USED IN THIS ANALYSIS.

1 SAS

55

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: ORGAN

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V
MODEL	2	463.57874844	231.78937422	232.12	0.0001	0.164670	34.001
ERROR	2355	2351.62735643	0.99856788		ROOT MSE		ORGAN MEA
CORRECTED TOTAL	2357	2815.20610487			0.99928368		2.9389313

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	463.57874844	232.12	0.0001

SAS

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN DATA SET = 2472

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2359

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	463.57874844	232.12	0.0001

SAS

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	3	A B C

NUMBER OF OBSERVATIONS IN DATA SET = 2472

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2359 OBSERVATIONS IN DATA SET CAN BE USED IN THIS ANALYSIS.

SAS

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: CONTENT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V
MODEL	2	530.69134329	265.34567164	250.48	0.0001	0.175349	33.434
ERROR	2356	2495.80039050	1.05933803		ROOT MSE		CONTENT MEA
CORRECTED TOTAL	2358	3026.49173379			1.02924148		3.0784230

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	2	530.69134329	250.48	0.0001

BEST COPY

1 SAS

TASK 7 ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	B C

NUMBER OF OBSERVATIONS IN DATA SET = 2186

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2059 OBSERVATIONS IN DATA SET CAN BE USED IN THIS ANALYSIS.

1 SAS 9:37 MONDAY, DECEMBER 2, 1985

TASK 7 ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: UI

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V
MODEL	1	46.32195254	46.32195254	46.46	0.0001	0.022090	31.329
ERROR	2057	2050.68921793	0.99693205		ROOT MSE		UI MEA
CORRECTED TOTAL	2058	2097.01117047			0.99846484		3.1869839

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	46.32195254	46.46	0.0001

1 SAS

TASK 7 ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	B C

NUMBER OF OBSERVATIONS IN DATA SET = 2186

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2041 OBSERVATIONS IN DATA SET CAN BE USED IN THIS ANALYSIS.

1 SAS

TASK 7 ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: COMP

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V
MODEL	1	38.75550781	38.75550781	57.05	0.0001	0.027220	26.014
ERROR	2039	1385.04513458	0.67927687		ROOT MSE		COMP MEA
CORRECTED TOTAL	2040	1423.80064239			0.82418243		3.1682181

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	38.75550781	57.05	0.0001

1 SAS 9:37 MONDAY, DECEMBER 2, 1985

BEST COPY

TASK 7 ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	B C

NUMBER OF OBSERVATIONS IN DATA SET = 2186

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2041

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	38.75550781	57.05	0.0001

1 SAS

TASK 7 ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	B C

NUMBER OF OBSERVATIONS IN DATA SET = 2186

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2041
OBSERVATIONS IN DATA SET CAN BE USED IN THIS ANALYSIS.

1 SAS 9:37 MONDAY, DECEMBER 2, 1985

TASK 7 ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: CONTENT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V
MODEL	1	43.94988807	43.94988807	45.90	0.0001	0.022013	30.807
ERROR	2039	1952.55182677	0.95760266		ROOT MSE		CONTENT MEA
CORRECTED TOTAL	2040	1996.50171485			0.97857175		3.1763841

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	43.94988807	45.90	0.0001

1 SAS

TASK 7 ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	B C

NUMBER OF OBSERVATIONS IN DATA SET = 2186

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2041
OBSERVATIONS IN DATA SET CAN BE USED IN THIS ANALYSIS.

SAS

BEST COPY

BEST COPY

SAS

TASK 7 ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: ORGAN

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V
MODEL	1	47.01021472	47.01021472	53.54	0.0001	0.025587	30.347
ERROR	2039	1790.29110817	0.87802408		ROOT MSE		ORGAN MEA
CORRECTED TOTAL	2040	1837.30132288			0.93702939		3.0877021

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	47.01021472	53.54	0.0001

1

SAS

TASK 7 ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	B C

NUMBER OF OBSERVATIONS IN DATA SET = 2186

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2041

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	47.01021472	53.54	0.0001

1

SAS

TASK 7 ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GRADE	2	B C

NUMBER OF OBSERVATIONS IN DATA SET = 2186

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2041
OBSERVATIONS IN DATA SET CAN BE USED IN THIS ANALYSIS.

1

SAS

TASK 7 ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: STYLE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V
MODEL	1	26.93927453	26.93927453	42.52	0.0001	0.020426	24.563
ERROR	2039	1291.94166619	0.63361533		ROOT MSE		STYLE MEA
CORRECTED TOTAL	2040	1318.88094072			0.79599958		3.2405683

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GRADE	1	26.93927453	42.52	0.0001

The significant differences for grade level support the sensitivity of the scoring rubrics to instructional exposure. Because the data will be reanalyzed by classroom level, distributed into strata by geography, etc., no inferences are made at this point. Preliminary correlational analyses were conducted to ascertain the relationship among scales. On inspection, these vary between .6 and .9. Our National Report will describe these analyses in detail. We also anticipate reporting the double scored reliabilities and the adjusted scores based upon calibration essay scoring for the raters. In addition, tasks 3, 5, and 9 will be analyzed.

Next Steps

Our next steps depend upon additional resources. We hope to secure support through OERI for the following efforts:

1. To create data files for student questionnaire and attitude data.
2. To create files for teacher and school level questionnaires.
3. To conduct descriptive analyses.
4. To conduct relational analyses based upon data in school and teacher questionnaires in order to identify promising instructional practices.
5. To prepare school level reports for local schools and states.
6. To prepare a report of the U.S. National Study.
7. To prepare files for international analysis of the USA data.

Were additional significant resources available, we would ideally like to score the remaining data from each classroom. We would also like to prepare qualitative analyses of the student papers, focusing on what they said, rather than mode of expression.

Summary

Our study has been successful to this point. We have achieved an appropriate response level and believe we have provided technically competent work within both time and budget constraints. We plan to share the results of our future analyses (when we have developed the full analyses) with researchers and practitioners to develop sound inferences about the quality of writing of American students.

SUPPORTING DOCUMENTS*

*Due to the cost of duplication, only one copy of supporting documentation is provided.

TASK 1 TRAINING AND SCORING MATERIALS

SCORING GUIDES

Introduction: You are to score a number of compositions written by a wide range of students from your country and to rate these compositions according to a scoring guide and a set of compositions whose scores have been determined. The guide and the set of compositions represent the consensus of experts from over fifteen countries as to the characteristics of a good composition on this task. You have probably taught a number of students like the ones whose compositions you will be scoring. In this case, however, you do not know the student personally and you have only these examples of their writing on which to judge the student's performance. As you score each composition, you should try to consider it in the light of the scores given the set of compositions that your group leader has provided; they will form a benchmark reference for your scores. It may be that you will not find a composition that reaches the peak of perfection that you would hope for from an experienced published writer; these are in most cases first drafts and you are asked to judge these compositions according to a consensus as to how well a student of this age or level of education can write in a limited period of time on a topic that is relatively unfamiliar to that student.

How you will score: A. You are first asked to rate each composition according to your overall impression of the quality of the composition. Your overall impression is usually based on your sense of how well the composition fulfills the task and topic requirements and your sense of its quality as a piece of writing relative to other students in the age-group that completed that task.

B. In addition you are asked to rate each composition as to its quality according to a number of specific dimensions: its content, its organization, its style and tone, its uses of grammar, orthographic conventions (i.e., spelling, punctuation), its neatness, and your personal reaction to the composition and its writer.

To help you make these judgements, we have had a set of compositions rated by a jury of experts in composition from several countries. They have determined that certain aspects of each composition need to be rated separately although they agree that an overall impression is also important.

The jury has developed a scoring scheme which you are asked to use, and they have created international guidelines and definitions of certain aspects within that scheme. The jury has also selected the benchmark compositions to illustrate various levels of performance by students with respect to the different parts of the scheme. Some of the compositions were written by students in other countries who were asked to write on the same topic. You are, therefore, asked to judge the compositions you will be reading according to an international scale.

You are going to spend several hours practicing using that scale. This guide is an introduction to the scale and the scheme of marking.

The Score Sheet. For each type of composition you will be given a score sheet. The sheet is designed so that you can indicate your ratings of ten compositions. A sample score sheet for Task Five appears on the next page. You must first enter the Student Number. Check to make sure you have the Student Number exactly. The Task Number has already been printed on the second row of the sheet.

Task 5 contains a number of choices as to the specific topic. Enter the Topic Number below the Task Number if you can determine which topic (1-7) the student selected. Then you must enter the number you have been assigned in each space on the row marked Rater Number. If you find that you cannot score the paper because the student wrote nothing or because the paper is illegible or for some other reason, indicate in the box in row 5/6 the reason for not scoring the composition using one of the codes 7-9, but DO NOT enter any scores beyond this point. If you do score the composition, indicate your Overall Impression in the box in row 5/6 and proceed down the column, entering a score for each aspect to be scored. DO NOT OMIT A SCORE.

Overall Impression. Read the composition in a normal fashion, without stopping to mark or underline segments of the composition or to write comments. Record your first impression of the overall quality of the composition. Indicate on a scale of 1 (Inadequate) to 5 (Excellent) what you think of the composition as a whole. DO NOT CHANGE THIS SCORE ONCE YOU HAVE RECORDED IT.

Rate the composition according to your overall impression of its merit. DO NOT attempt to rate your estimate of where it would fall on a normal curve of performance, but your sense of its adequacy or excellence as a composition according to the benchmark scores. Thus, you should not award a certain proportion of low, intermediate, or high grades, if you do not perceive such a distribution in the set of compositions you are asked to score.

Detailed Impression. After you give your overall impression of the composition, you will also rate each composition on a number of detailed dimensions, which are slightly modified or expanded according to the nature of the writing task. These dimensions represent an international consensus of teachers and judges of writing about those aspects of a composition that should be singled out for attention. Some dimensions may be more important in one country; other dimensions in another. We are seeking to have each country's students judged according to an international set of standards, so you should consider each dimension equally. You do not have to try to make your rating of each dimension match or average to equal your overall impression.

AFTER YOU RATE ALL THE DIMENSIONS, DO NOT CHANGE YOUR OVERALL IMPRESSION SCORE.

The dimensions that will be included are:

QUALITY AND SCOPE OF IDEAS. A dimension that focuses on your impression of the content of the composition, what it says and how fully and completely it says it.

ORGANIZATION AND PRESENTATION OF CONTENT. How the student arranges the material and the structure of both the composition as a whole and the individual segments or paragraphs.

STYLE AND TONE. Including the choice of words and phrases, sentence structures and larger units of discourse, the variety and flow of sentences, and your sense of whether the student has made an effective and appropriate use of the language given the aim, context, and audience of the composition task.

SAMPLE SCORING SHEET TASK 5

1. Student Number										
2. Task Number	15	15	15	15	15	15	15	15	15	15
3. Topic Number (1-7)										
4. Rater Number (your number)										
5. Composition Cannot Be Scored (7-9)*										
6. Overall Impression (1-5)										
7. Quality and Scope (1-5)										
8. Thematic Appropriateness (1-5)										
9. Presentation of Characters (1-5)										
10. Organization and Presentation (1-5)										
11. Overall Structure (1-5)										
12. Control of Detail (1-5)										
13. Style and Tone (1-5)										
14. Choice and Consistency of Tone (1-5)										
15. Choice of Words and Phrases (1-5)										
16. Grammatical Features (1-5)										
17. Spelling and Orthography (1-5)										
18. Handwriting (1-5)										
19. Interest in the Composition (1-5)										
20. Sense of Connection with Writer (1-5)										

*If the composition cannot be scored, mark in row 5 one of the following: 7 = off topic
8 = illegible
9 = blank paper, no response

LEXICAL AND GRAMMATICAL FEATURES. Including both lexical and syntactic features to indicate your impression of the student's mastery of these features.

SPELLING AND ORTHOGRAPHIC CONVENTIONS. Including those conventional aspects of punctuation to indicate your impression of the student's mastery of these features.

HANDWRITING. To indicate your sense of the student's physical presentation of the composition including legibility and clarity of corrections and paragraph signals. You should remember that a student may not have had time to recopy the whole composition.

RESPONSE OF THE RATER. To provide you with an opportunity to indicate your interest in the composition, and your intuitive sense of the writer as seen through the composition, whether you are attracted to the person or are persuaded by what the person has written.

You should rate each composition following the order of these dimensions that have been outlined on the scoring sheet. MARK EACH COMPOSITION ON EACH SCORE CALLED FOR.

Sub-category Ratings. On Tasks 5, 6 and 7, which are more complex, you will be asked to rate the composition on certain particular aspects of one or more of the dimensions. Again these are aspects that the international jury has agreed should be rated independently. You should rate the category itself first and then the sub-categories. The category score is not an average of the sub-categories for it includes more than their sum. There will be detailed explanations of these aspects in the scoring guide for the particular task.

Benchmark Compositions. Attached to the scoring guide for each task is a set of compositions that have been selected by the international jury to help you understand the interpretation of the scoring guide and its application to specific compositions. As you read through the scoring guide, you should also read through the benchmark compositions, and the associated commentaries. Doing so will help you become familiar both with the dimensions that will be rated and with the international standards that set the scale from Excellent to Inadequate on each of the categories and sub-categories. While scoring compositions after the training session, you should frequently refer to the benchmark compositions to help you to apply the international standard.

BEST COPY

TASK 1A: Description of a Bicycle

A. QUALITY AND SCOPE OF CONTENT. In making this judgement, you should look for:

1. whether the student has provided sufficient characteristics (e.g., color, type -- racing, touring, boy's or girl's -- and accessories) so that the uncle could purchase the bicycle requested, and;

2. whether the student has omitted superfluous details (e.g., explaining that a wheel has spokes). (In some cases the student has described a bicycle other than those pictured. Unless forbidden in a country's directions to the student, this response is acceptable.)

B. ORGANIZATION AND PRESENTATION OF CONTENT. In making this judgement, you should determine whether the characteristics are grouped naturally into a structure that makes sense to a reader seeking to identify a bicycle. The writer should not make many false starts.

C. STYLE AND TONE. In making this judgement, you should consider:

1. the selection of appropriate terms to describe the features of the bicycle.

2. the extent to which the specific words, phrases, and larger units of discourse show that the writer is aware that the audience of the letter is a generous older person. You should be aware that the stimulus contains the opening sentence so that a good letter may appear to begin and end abruptly.

D. LEXICAL AND GRAMMATICAL FEATURES. To be supplied by National Centers.

E. SPELLING AND ORTHOGRAPHIC CONVENTIONS. To be supplied by National Centers.

F. HANDWRITING AND NEATNESS. To be supplied by National Centers.

G. RESPONSE OF RATER. You should give an estimate of your interest in the composition and/or the writer.

TASK 1B: Self Description

A. QUALITY AND SCOPE OF CONTENT. In making this judgement you should consider the use of essential and relevant information including sufficient observable characteristics, and distinctive characteristics concerning clothes, the place where the writer will be, or other information appropriate to the situation.

B. ORGANIZATION AND PRESENTATION OF CONTENT. In making this judgement you should consider:

1. the use of continuous discourse (not a list of features);

2. the orderly presentation of information to provide a clear description.

C. STYLE AND TONE. In making this judgement you should consider:

1. the selection of descriptive terms.

2. the appropriate choice of words, phrases and larger units of discourse to indicate the writer's awareness of the individual who is the audience (the writer has the option to be casual and informal or relatively formal).

D. LEXICAL AND GRAMMATICAL FEATURES. To be supplied by National Centers.

E. SPELLING AND ORTHOGRAPHIC CONVENTIONS. To be supplied by National Centers.

F. HANDWRITING AND NEATNESS. To be supplied by National Centers.

G. RESPONSE OF RATER. You should give an estimate of your interest in the composition and/or the writer.

TASK 1C: Letter to Headmaster

A. QUALITY AND SCOPE OF CONTENT. In making this judgement, you should consider that there are 5 relevant facts we might expect in this message: 1) mentioning of appointment, 2) mentioning of fact that the writer is hindered from coming, 3) mentioning of reason for non-appearance, 4) apology, 5) name of student.

B. ORGANIZATION AND PRESENTATION OF CONTENT. In making this judgement, you should consider that there is not one single order of giving the four elements mentioned under A. You should decide if the content is presented in an orderly manner.

C. STYLE AND TONE. In making this judgement, you should consider:

1. the tone of the message should be in accordance with the relationship between headteacher (or principal) and student;

2. the style should show an appropriate use of words, phrases, and larger units of discourse.

D. LEXICAL AND GRAMMATICAL FEATURES. To be supplied by National Centers.

E. SPELLING AND ORTHOGRAPHIC CONVENTIONS. To be supplied by National Centers.

F. HANDWRITING AND NEATNESS. To be supplied by National Centers.

G. RESPONSE OF RATER. You should give an estimate of your interest in the composition and/or the writer.

TASK 1D (Population A): Informal Note to Family

A. QUALITY AND SCOPE OF CONTENT. In making this judgement, you should consider that the provision of essential and relevant information includes the addressee, details of where the writer can be found, with whom the writer will be and how long the writer will remain. The writer should give his/her name. Both a bare listing of information and a more elaborate presentation are acceptable.

B. ORGANIZATION AND PRESENTATION OF CONTENT. In making this judgement, you should consider the clear and orderly presentation of information. The message may be short, but the information must be presented in an order so that the reader can gather sufficient detail.

C. STYLE AND TONE. In making this judgement, you should consider the choice of words, phrases, and larger units of discourse including the use of modes of address and other words appropriate for an informal message to a member of one's own family.

D. LEXICAL AND GRAMMATICAL FEATURES. To be supplied by National Centers.

E. SPELLING AND ORTHOGRAPHIC CONVENTIONS. To be supplied by National Centers.

F. HANDWRITING AND NEATNESS. To be supplied by National Centers.

G. RESPONSE OF RATER. You should give an estimate of your interest in the composition and/or the writer.

TASK 1E (Population B and C): Application for a Holiday Job

A. QUALITY AND SCOPE OF CONTENT. In making this judgement, you should consider:

1. whether the letter clearly refers to the job for which the candidate is applying;

2. specifies the period when the candidate will be available;

3. indicates that the applicant is qualified for the position by age or educational level;

4. gives the name and address and/or phone number (although in some countries that information can be on the envelope, it might well be referred to in the text of the letter).

B. ORGANIZATION AND PRESENTATION OF CONTENT. In making this judgement, you should consider:

1. the clear and orderly presentation of information including the use of appropriate format for a letter of application;

2. the orderly presentation of information called for in the advertisement.

C. STYLE AND TONE. In making this judgement, you should consider that the choice of words, phrases, and larger units of discourse (such as modes of address and salutation) are appropriate to a formal letter of inquiry. You may wish to assume the role of the recipient and decide whether you would be favorably inclined towards the writer.

D. LEXICAL AND GRAMMATICAL FEATURES. To be supplied by National Centers.

E. SPELLING AND ORTHOGRAPHIC CONVENTIONS. To be supplied by National Centers.

F. HANDWRITING AND NEATNESS. To be supplied by National Centers.

G. RESPONSE OF RATER. You should give an estimate of your interest in the composition and/or the writer, including the degree to which you might respond favorably to the writer.

Explicit Features Used in Task 1 Training and Scoring

TASK 1A	CONTENT	ORGANIZATION	STYLE TONE
	Unique characteristics, type, boys/girls, 10 speed?, handlebars, baskets, trainingwheels, horns, pockets, lights?		
5	All items & comments or function, desirability & recognition of reader	Must be cohesive & coherent has intro, body, & conclusion paraphrasing, follow a progression a plan in description eg. top to bottom or general to specific; no digression or lapses in logic	Effective use of descriptions, sentence variety, acknowledging reader politely, but no obsequiously, shows gratitude
4	All items there but not all the detail	Includes all in 5 but plan not as clear - must have cohesion & coherence - does not require paragraphing	All of 5 but used inconsistently
3	Missing detail, but not enough to lose identification of the bicycle - some superflures detail ok	No diliberate plan, however organization does not impede comprehension	Lack of descriptors, and sentence variety - little acknowledgement of reader weak diction
2	Can't tell I.D. of bike, vague	Confusion, no discernable plan	Somewhat like a list, not obvious,polite, or grateful
1	Can't tell I.D. of bike - have not fulfilled task, missing information	Incomprehensible	Rude, ie demanding, no acknowledgement of reader

TASK 1B	CONTENT	ORGANIZATION	STYLE TONE
	Observable distinctive characteristrics, such as physical characteristics, size, shape, hair, coloring, age, sex, clothing and other people or props; may mention a special or specific meeting place		
5	Can easily identify person includes several distinctive characteristics	Must be cohesive & coherent has intro, body, & conclusion paraphrasing, follow a progression a plan in description eg. top to bottom or general to specific; no digression or lapses in logic	Effective use of descriptions, sentence variety, acknowledging reader warmly in friendly manner, but no obsequiously, shows gratitude
4	Can identify, has some distinctive characteristics	Includes all in 5 but plan not as clear - must have cohesion & coherence - does not require paragraphing	All of 5 but used inconsistently
3	Probably can but might have some difficulty; too many common characteristics but not enough unique	d diliberate plan, however organization does not impede comprehension	Lack of descriptors, and. sentence variety - little acknowledgement of reader weak diction
2	Can't I.D. the person;includes some irrelevant features	Confusion, no discernable plan	Somewhat like a list
1	Includes mostly irrelevant, unobservable features; can't identify person	Incomprehensible	Rude, no acknowledgement of the reader

TASK 1C	CONTENT	ORGANIZATION	STYLE TONE
	Reason for nonappearance, apology, name, consideration for principal's time		
5	All of the above items with some details, explanation and perhaps alternative suggestion	Correct format: greeting, body, closing; cohesive; "paragraphing" when appropriate; no digression	Formal, courteous, recognition authority of principal message sounds sincere and the reason is important, choice of vocabulary, sentence variety
4	All items there but not as elaborate	correct format, minor problems in cohesion, a small digression	All of 5 but less variety in sentence structure
3	Missing one key item (bare bones) adequate	Correct format, very short eg: 1-2 short sentences, does not impede reader understanding	Somewhat polite, courteous; little acknowledgement of reader - lacks sentence variety
2	Missing two or more	flawed format, impedes reader understanding	Somewhat list-like not obviously polite poor diction
1	Declaration of absence or excuse	No format, confusing, digression	No acknowledgement of reader, rude, telegraphic

TASK 1D	CONTENT	ORGANIZATION	STYLE/TONE
	Addressee, where writer is, with whom, how long grd., when will return, some indication of who writer is	"P.S." - If used, P.S. should not have essential information for task.	
5	All of above, plus elaboration of some, such as show of concern, phone #, chaperones, safety	Has sufficient content to have introduction, body and conclusion, coherence and cohesion effective	Shows concern for reader, i.e., personal information; established tone of safety and well-being; sentence variety
4	Contains everything but not as elaborate or detailed	Has evidence of introduction, body and conclusion, but not necessarily all three; coherence and cohesion effective	All of 5 but used inconsistently
3	Missing one element, "bare bones"	Adequate, "bare bones," but too short; minor problems in cohesion	Neutral in tone, lacks feeling or concern; lack of sentence variety
2	Missing two or more elements; lack of concern for reader	Real problem in coherence or cohesion; impedes reader's understanding	Little or no attempt at sentence variety; shows little or no concern for parent; somewhat like a list; poor diction
1	Declaration of absence; no concern for reader	No format; no evidence of a plan or organization; no logical relationships	Abrupt

TASK 1E	CONTENT	ORGANIZATION	STYLE/TONE
	↓	↓	↓
5	Name, address, and/or phone # of writer; reference to the job; period candidate is available; qualifications; examples of age, experience; availability for interview	Must have appropriate formal letter style; inside and return address; salutation; body; signature/closure; cohesive and coherent transitions	Formal letter style recognizes status of reader; business-like sentence variety and diction; writer sells himself to reader
4	Has all but does not elaborate, or less elaborately done	Has all or most of the elements of the "5" paper; must have either an address or a phone #; minor problem in cohesion, e.g., a small digression	Not as successful in selling himself - not as business-like
3	Lacks one element; little or no elaboration	Flawed format, missing element; minor problem with coherence; organization does not indicate reader's understanding	Little sentence variety; inconsistent in tone
2	Lacks more than one element	Little or no evidence of format; missing two or more elements; organization impedes reader's understanding	No sentence variety; too casual
1	Missing most elements	No format; confusing; digressions	No attempt to vary structure; rude; demanding

BEST COPY

FULL ROTATION OF TASKS FOR GRADES 6, 10, AND 12